Longitudinal MRI Data Analysis: Mixed Method ANOVA

I. Introduction

This report explores the data from a longitudinal study on MRI results of patients with/without dementia. The analysis is guided by the research questions formulated as a result of the Exploratory Data Analysis (EDA). The datasets include cross-sectional and longitudinal MRI data from young, middle-aged, nondemented, and demented older adults.

II. Data Cleaning and Wrangling

The dataframe has a total of 294 rows and 16 columns with no duplicated values. The dataframe has some missing values for the columns “SES” missing 15 variables, and variable “MMSE” has one missing value, which is dropped. The new “Agegroup” categorical variable has been created based on the continuous variable “Age”. A new variable subject has been created based on the “Subject ID” to indicate unique cases.

For this analysis, the following variables have been selected: "Group", "Visit", "MMSE", "Age", "Agegroup", "eTIV", "nWBV", "ASF".

III. Exploratory Data Analysis

The analysis extensively explored the dataset, and it has been reflected in the Python notebook(‘NARGIZ-GULIYEVA-A4’). Only key results have been included in this paper.

Based on the Table 1, a cohort of individuals between age 60 and 98 years old has been examined in this dataset. The average age of the individuals of 76.4 years. A mean score of cognitive function measured by Mini-Mental State Examination (MMSE) is 27.3, which shows average intact cognitive abilities with variability reflected in standard deviation 3.4. The estimated total intracranial volume (eTIV) exhibited a mean value of 1478.9, with a standard deviation of 176.6. The normalized whole brain volume (nWBV) showed relatively little variability, with a mean of 0.7. The atlas scaling factor (ASF), a measure used to normalize brain volumes in MRI analysis, displayed a mean of 1.2, with a standard deviation of 0.1.

	Age	MMSE	eTIV	nWBV	ASF
mean	76.4	27.3	1478.9	0.7	1.2
std	7.6	3.4	176.6	0.0	0.1
min	60	15	1106	0.6	0.9
max	98	30	2004	0.8	1.6

Table 1

Figure 1 shows that median MMSE score and its variability between visits 1 and 2 are quite close. The boxplot on Figure 2 shows that eTIV score differs by the age group. Based on the Figure 2, age group 2 stands out as having the highest central tendency and the highest variability, while age group 4 has the lowest median with less variability.

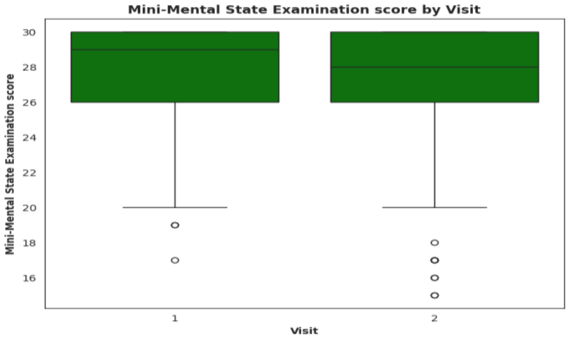

Figure 1

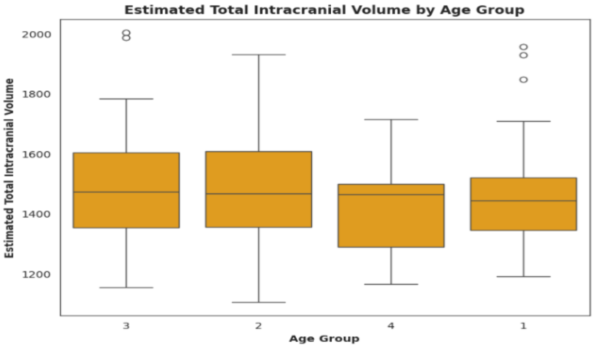

Figure 2

According to the Figure 3, the mean MMSE scores across visits for the converted group are higher than those for the demented group, indicating potentially better cognitive function in the converted group. Likewise, the mean MMSE scores for the nondemented group are higher than those for the demented group. The variability of visits within the demented and converted groups are high. This examination has driven the further analysis of the research question of the effect of cognitive status and visits on MMSE scores.

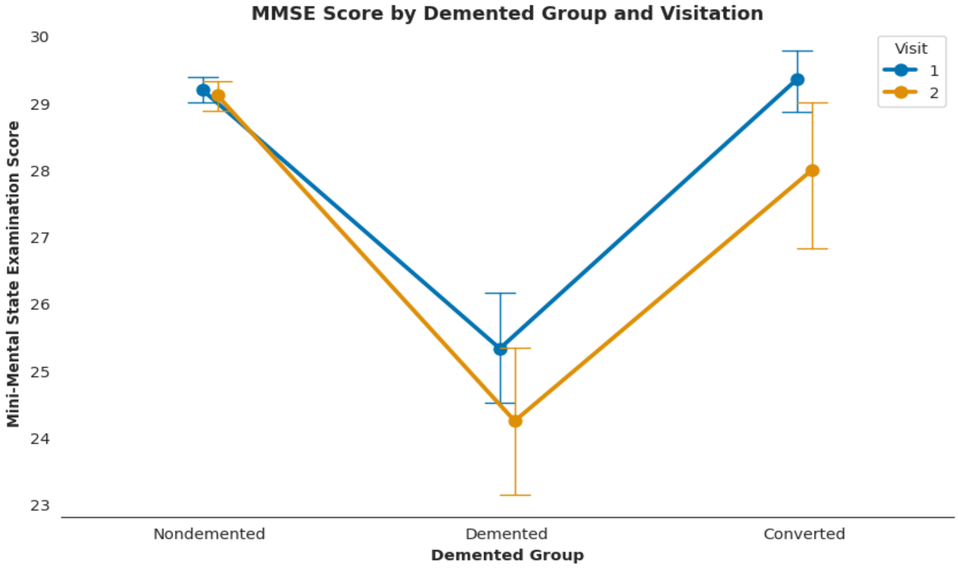

Figure 3

According to Figure 2 that compares means, there are noticeable differences in the in the mean eTIV between visits. The patients between 70 and 79 years old (group 3) exhibit the highest mean eTIV values across both visits, the patients between 90 and 99 years old (group 4) consistently demonstrate the lowest mean eTIV values. Within each age group, there are variations in the mean eTIV values between initial and subsequent visits. This has further analyzed as a research question of the impact of age groups and visits on eTIV.

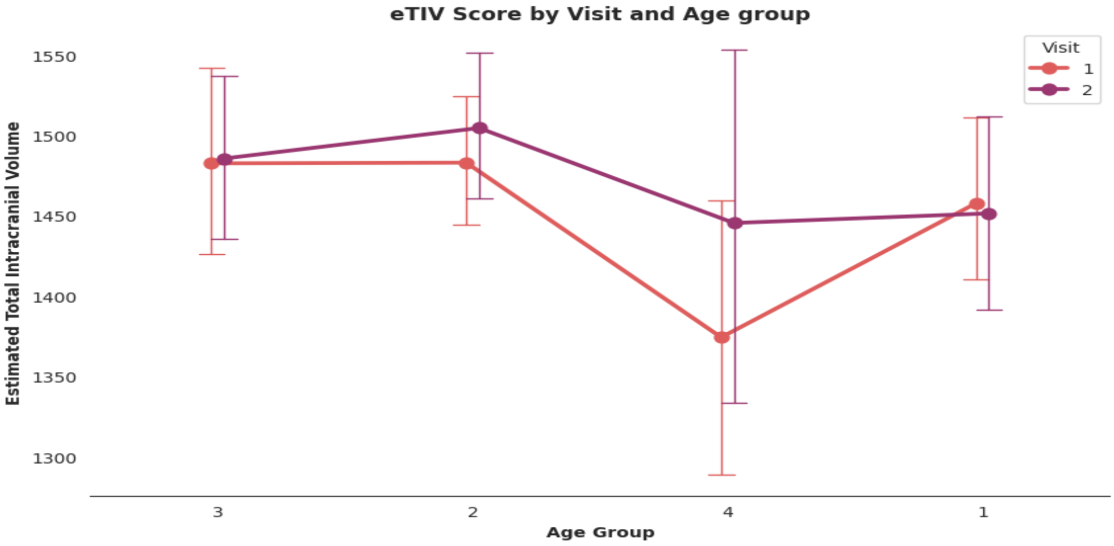

Figure 4

IV. Model 1: Cognitive Status and Visit Sequence Effects on MMSE Scores

Based on the EDA, the first **research question** is: what is the effect of cognitive status and initial or subsequent visits on MMSE scores among individuals aged 60 and above? The mixed-method ANOVA has been employed to answer the research question.

The analysis indicates that both cognitive status and visits have significant effects on MMSE scores among individuals aged 60 and above, and there is also a significant interaction effect between cognitive status and visit on MMSE scores. Based on the results, the main effect of Group is statistically significant ($p < 0.001$), rejecting the null hypothesis. Furthermore, the main effect of Visit is statistically significant ($p = 0.003$), rejecting the null hypothesis as well. The interaction term between Group and Visit is statistically significant ($p = 0.037$), rejecting the null hypothesis.

The assumption tests have been utilized to evaluate the model, and those suggest the following:

- **Assumption 1: Sphericity.** The value of 1.0 indicates perfect sphericity, so the variances of the differences between all pairs of within-subject conditions are equal.

- **Assumption 2: Normality.** The MMSE scores for all groups do not follow a normal distribution based on the Shapiro-Wilk test, as indicated by the significant p-values. Based on Shapiro-Wilk statistic, p-value equals 2.64e-12 for the nondemented group; p-value equals to 5.68e-06 for the Demented group; and the p-value is 5.87e-05.

- **Assumption 3: Homogeneity.** The MMSE scores for both Visit 1 and Visit 2 exhibit significant violations of the homogeneity of variance assumption across the nondemented, demented, and converted groups. The Levene's test shows associated p-value equals to 1.04e-13 for visit 1. Additionally, based on the Levene's test for visit 2, p-value is 1.13e-12.

Assumption on normality and homogeneity have been violated, which affect the validity and reliability of the model. Taking this into consideration, alternative methods need to be considered.

V. Model 2: Age Group and Visit Effects on eTIV

The next **research question** is: what is the impact of age group and initial or subsequent visit on eTIV among individuals aged 60 and above? The mixed method ANOVA has been employed in this instance as well.

Based on the results, while there are significant differences in eTIV between initial and subsequent visits, there are no significant differences in eTIV among individuals belonging to different age groups. Additionally, there is no significant interaction between age group and visit on eTIV. The main effect of age groups is not statistically significant ($p = 0.483$). The main effect of visit is statistically significant ($p = 0.035$). The interaction effect between age groups and visit is not statistically significant ($p = 0.739$).

The assumption tests have been utilized to evaluate the model, and those suggest the following:

- **Assumption 1: Sphericity.** The result of 1.0 indicates perfect sphericity, so there is no violation of this assumption in this model.

- **Assumption 1: Normality.** Based on the Shapiro-Wilk test, the distributions of eTIV within age groups 1, 2, and 3 significantly deviate from normality, while the distribution within age group 4 does not. The associated p-value is 0.00014 for age group 1; the p-value is approximately 0.01035 for age group 2; the p-value is approximately 0.04187 for age groups 3; the p-value is approximately 0.422 for age groups.

- **Assumption 3: Homogeneity.** Based on the Levene test, the results indicate that the assumption of homogeneity of variances is met for both the first and second visits, as the variances of eTIV are approximately equal across different age groups within each visit. The associated p-value equals 0.35 for visit 1, p-value is approximately 0.66 for visit 2.

Only normality assumption is violated. This might affect the validity of statistical analyses, and alternative methods may be required for robust data analysis.

VI. Sample size and power analysis

The result of the power analysis indicates that a sample size of approximately 45 individuals per group would be needed to achieve a power of 0.91 with a significance level of 0.05 and assuming an effect size of 0.7.

The power analysis plot displays insight into the relationship between sample size and statistical power (Figure 5). The dashed red line at the sample size of 45 indicates where the specified sample size falls on the plot helping to visualize the power associated with a given sample size.

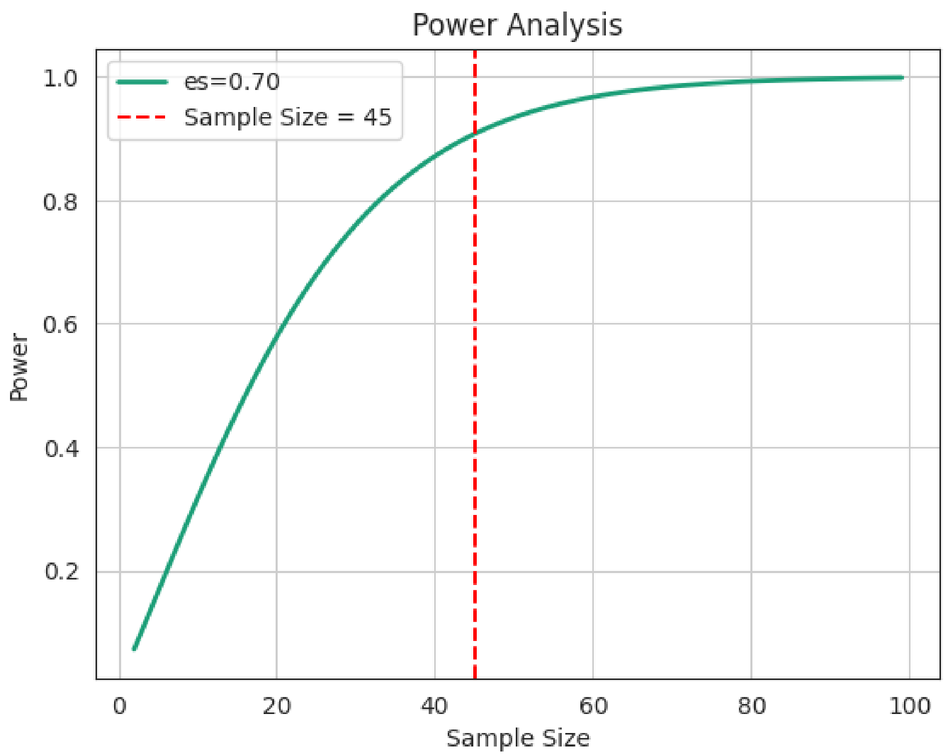

Figure 5